4.3 Autour de la décomposition en valeurs singulières (148, 154, 157, 162, 219 (?)) [2] [26]

La décomposition en valeurs singulières est un outil intéressant pour le traitement de données. En effet, elle apparaît naturellement dans la résolution de problèmes au sens des moindres carrés (et donc en régression) mais aussi dans des outils de compression de données (cela se verra dans les résultats que nous allons énoncer) ! Le but est donc de montrer le résultat suivant :

Théorème 4.5 (Décomposition en valeurs singulières (aka SVD))**.** Soient $m, n \in \mathbb{N}^*$ et $A \in \mathscr{M}_{m,n}(\mathbb{R})$ (resp. $\mathscr{M}_{m,n}(\mathbb{C})$). Alors il existe $(U, V) \in \mathrm{O}_m(\mathbb{R}) \times \mathrm{O}_n(\mathbb{R})$ (resp. $\mathrm{U}_m(\mathbb{C}) \times \mathrm{U}_n(\mathbb{C})$) et il existe d'uniques réels $\sigma_1 \geqslant \ldots \geqslant \sigma_r > 0$ (r étant le rang de A) tels que A s'écrive :

$$A = U\Sigma V^T \quad (\text{resp.} \quad A = U\Sigma V^*)$$

où $\Sigma \in \mathscr{M}_{m,n}(\mathbb{R})$ désigne la matrice :

$$\Sigma := \begin{pmatrix} \begin{matrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_r \end{matrix} & \mathbf{O}_{r,n-r} \\ \mathbf{O}_{m-r,r} & \mathbf{O}_{m-r,n-r} \end{pmatrix}.$$

On appelle cette décomposition de A la *décomposition en valeurs singulières* de A, et les réels σ_i sont appelés *valeurs singulières* de A.

Cette décomposition permet également de montrer les résultats suivants (dans la suite, $\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$ et la notation * unifie la transposée et la transconjuguée) :

Théorème 4.6 (Eckart-Young)**.** Soit $A \in \mathscr{M}_{m,n}(\mathbb{K})$ de SVD $U\Sigma V^*$ et de rang r. Alors, en notant, pour $k \in [\![1, r]\!]$:

$$A_k := U\Sigma_k V^*, \quad \text{où} \quad \Sigma_k := \begin{pmatrix} \begin{matrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_k \end{matrix} & \mathbf{O}_{k,n-k} \\ \mathbf{O}_{m-k,k} & \mathbf{O}_{m-k,n-k} \end{pmatrix},$$

alors A_k réalise le minimum suivant :

$$|\!|\!| A - A_k |\!|\!|_2 = \min_{\substack{B \in \mathscr{M}_{m,n}(\mathbb{K}) \\ \mathrm{rg}(B) = k}} |\!|\!| A - B |\!|\!|_2.$$

Remarque 4.3.1. *C'est* **ce** *théorème qui justifie le plus, selon moi, le fait de placer ce développement potentiel dans la leçon 148 : c'est indiqué dans le rapport du jury, tout comme le problème des moindres carrés qui arrive juste après ! De plus, on utilise dans la preuve de ce théorème un argument de dimension pour justifier que l'intersection de deux sous-espaces est non-nulle.*

Théorème 4.7 (Résolution au sens des moindres carrés)**.** Soit $A \in \mathscr{M}_{m,n}(\mathbb{K})$. Notons $U\Sigma V^*$ sa SVD, et notons :

$$A^\dagger := V\Sigma^\dagger U^* \in \mathscr{M}_{n,m}(\mathbb{K}), \quad \text{où} \quad \Sigma^\dagger := \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 & \\ 0 & \frac{1}{\sigma_2} & \ddots & \vdots & \mathbf{O}_{r,m-r} \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \cdots & 0 & \frac{1}{\sigma_r} & \\ & & \mathbf{O}_{n-r,r} & & \mathbf{O}_{n-r,m-r} \end{pmatrix} \in \mathscr{M}_{n,m}(\mathbb{K}).$$

Cette matrice est souvent appelée *pseudo-inverse* de la matrice A. On a alors, pour $b \in \mathbb{K}^m$, que le vecteur $x_b := A^\dagger b \in \mathbb{K}^n$ est une solution de l'équation $Ax = b$ au sens des moindres carrés, c'est-à-dire :

$$\|Ax_b - b\|_2 = \min_{x \in \mathbb{K}^m} \|Ax - b\|_2$$

et cette solution est l'unique solution de norme minimale, c'est-à-dire que pour tout $x \in \mathbb{K}^n$ solution de $Ax = b$ au sens des moindres carrés, si $x \neq x_b$, alors :

$$\|x_b\|_2 < \|x\|_2.$$

Démonstration de la SVD. On se place dans $\mathbb{K} = \mathbb{R}$, la preuve étant exactement la même pour $\mathbb{K} = \mathbb{C}$. Soit $M := A^T A$. On a alors :

1. $M \in \mathscr{S}_n^+(\mathbb{R})$,
2. $\ker(M) = \ker(A)$, et donc, en particulier : $\mathrm{rg}(M) = \mathrm{rg}(A) := r$.

Il existe alors $V \in \mathrm{O}_n(\mathbb{R})$ et $\lambda_1 \geqslant \ldots \lambda_r > 0 = \lambda_{r+1} = \ldots = \lambda_n$ tels que :

$$M = A^T A = V \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) V^T.$$

Étant donné que $\lambda_1 \geqslant \ldots \geqslant \lambda_r > 0$, on peut alors trouver $\sigma_1 \geqslant \ldots \geqslant \sigma_r > 0$ tels que :

$$\forall i \in [\![1, r]\!], \quad \sigma_i^2 = \lambda_i.$$

Posons alors $\Sigma := \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \ddots & \vdots & \mathbf{O}_{r,n-r} \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \cdots & 0 & \sigma_r & \\ & & \mathbf{O}_{m-r,r} & & \mathbf{O}_{m-r,n-r} \end{pmatrix} \in \mathscr{M}_{m,n}(\mathbb{R})$. Le but est alors de trouver une matrice $U \in \mathrm{O}_m(\mathbb{R})$ telle que :

$$AV = U\Sigma,$$

c'est-à-dire, en notant v_i les colonnes de V et u_i les colonnes de U, trouver une base orthonormée $(u_1, \ldots, u_m)$ de $\mathbb{R}^m$ telle que :

$$\forall i \in [\![1, r]\!], \quad Av_i = \sigma_i u_i \quad \text{et} \quad \forall i \in [\![r+1, n]\!], \quad Av_i = 0. \tag{4}$$

La deuxième condition est d'ores et déjà vérifiée. En effet, on a que, pour $i \in [\![r+1, n]\!]$, $v_i \in \ker(M) = \ker(A)$. Pour trouver les u_i, on ne va pas se casser la tête ! Définissons :

$$\forall i \in [\![1, r]\!], \quad u_i := \frac{1}{\sigma_i} A v_i \in \mathbb{R}^m. \tag{5}$$

Montrons que la famille $(u_1, \ldots, u_r)$ est une famille orthonormée de $\mathbb{R}^m$:

$$\forall (i,j) \in [\![1, r]\!]^2, \quad \langle u_i, u_j \rangle = \frac{1}{\sigma_i \sigma_j} \langle A v_i, A v_j \rangle = \frac{1}{\sigma_i \sigma_j} \left\langle v_i, A^T A v_j \right\rangle = \frac{1}{\sigma_i \sigma_j} \langle v_i, \lambda_j v_j \rangle = \frac{\sigma_j^2}{\sigma_i \sigma_j} \langle v_i, v_j \rangle = \delta_{i,j}.$$

Dans ce calcul, on a utilisé le fait que la famille $(v_1, \ldots, v_r)$ était une famille orthonormée de vecteurs propres de $A^T A$, pour les valeurs propres $\lambda_1, \ldots \lambda_r$ qui sont égales à $\sigma_1^2, \ldots, \sigma_r^2$. Ainsi, la famille $(u_1, \ldots, u_r)$ est une famille orthonormale de $\mathbb{R}^m$. On peut donc la compléter en une base orthonormale $(u_1, \ldots, u_m)$ de $\mathbb{R}^m$. En notant alors U la matrice dont les colonnes sont formées des vecteurs u_i, on a bien que $U \in \mathrm{O}_m(\mathbb{R})$ et que, par définition des vecteurs u_i en (5), on retrouve l'expression (4), qui est équivalente à :

$$AV = U\Sigma, \quad \text{i.e.} \quad A = U \Sigma V^T.$$

Il reste à dire que la matrice Σ est bien unique, étant donné qu'elle est entièrement déterminée par le rang de A et par les valeurs propres rangées dans l'ordre décroissant de $A^T A$. Cela conclut donc la preuve de la décomposition en valeurs singulières ! $\square$

Démonstration du théorème d'Eckart-Young. Soit $A \in \mathscr{M}_{m,n}(\mathbb{K})$, de SVD $U \Sigma V^*$. Pour toute matrice $B \in \mathscr{M}_{m,n}(\mathbb{K})$, on a :

$$|\!|\!| A - B |\!|\!|_2 = |\!|\!| U \Sigma V^* - B |\!|\!|_2 = |\!|\!| U \left(\Sigma - U^* B V \right) V^* |\!|\!|_2 = |\!|\!| \Sigma - U^* B V |\!|\!|_2$$

étant donné que les matrices U et V sont unitaires (ou orthogonales). Par ce même argument, on a en particulier que U et V sont inversibles, et donc, que $A_k \in \mathscr{M}_{m,n}(\mathbb{K})$, de rang k minimise $B \mapsto |\!|\!| A - B |\!|\!|_2$ si et seulement si $A'_k := U^* A_k V$, de rang k également, minimise $B' \mapsto |\!|\!| \Sigma - B' |\!|\!|_2$. Soit alors $B' \in \mathscr{M}_{m,n}(\mathbb{K})$ quelconque, de rang k. Notons $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{K}^n$, et considérons le sous-espace $\mathrm{Vect}(e_1, \ldots, e_{k+1})$. Par théorème du rang, on sait que la dimension de $\ker(B')$ est $n - k$. Ainsi :

$$\ker(B') \cap \mathrm{Vect}(e_1, \ldots, e_{k+1}) \neq \{0\}.$$

Prenons alors x de norme 1 dans cette intersection. Il s'écrit alors $x = \displaystyle\sum_{i=1}^{k+1} x_i e_i$ et on a donc :

$$(\Sigma - B')x = \Sigma x - 0 = \sum_{i=1}^{k+1} x_i \sigma_i e_i.$$

Ainsi :

$$|\!|\!| \Sigma - B' |\!|\!|_2 \geqslant \|(\Sigma - B')x\|_2 = \sqrt{\sum_{i=1}^{k+1} \underbrace{\sigma_i^2}_{\geqslant \sigma_{k+1}^2} |x_i|^2} \geqslant \sigma_{k+1}.$$

En prenant $A'_k = \Sigma_k$ comme définie dans l'énoncé, on a alors :

$$\forall x \in \mathbb{K}^n, \quad \|(\Sigma - \Sigma_k)x\|_2^2 = \sum_{i=k+1}^{r} \sigma_i^2 |x_i|^2 \leqslant \sigma_{k+1}^2 \sum_{i=k+1}^{r} |x_i|^2 \leqslant \sigma_{k+1}^2 \|x\|_2^2.$$

Par définition de la norme d'opérateur, on obtient donc :

$$|\!|\!| \Sigma - \Sigma_k |\!|\!|_2 \leqslant \sigma_{k+1}.$$

Ainsi, Σ_k réalise bien le minimum de la fonction $B' \mapsto \|\Sigma - B'\|_2$ sur les matrices de rang k, et donc, la matrice $A_k = U\Sigma_k V^*$ réalise le minimum de la fonction $B \mapsto \|A - B\|_2$ sur les matrices de rang k, ce qui conclut la preuve ! $\square$

Démonstration du théorème 4.7. **Étape 1 : L'équation normale, une condition nécessaire et suffisante**
Montrons qu'un vecteur $x \in \mathbb{K}^n$ est solution du système $Ax = b$ au sens des moindres carrés si et seulement si x est solution de l'équation normale :

$$A^*Ax = A^*b.$$

Pour cela, on pose $J : x \mapsto \|Ax - b\|_2^2$. On a alors que J est de classe $\mathscr{C}^\infty$ sur $\mathbb{K}^n$, et :

$$\forall x \in \mathbb{K}^n, \quad J(x) = \langle Ax - b, Ax - b\rangle = \langle Ax, Ax\rangle - 2\langle Ax, b\rangle + \langle b, b\rangle = \langle x, A^*Ax\rangle - 2\,\langle x, A^*b\rangle + \langle b, b\rangle.$$

On trouve donc :

$$\forall x \in \mathbb{K}^n, \quad \nabla J(x) = 2\left(A^*Ax - A^*b\right) \quad \text{et} \quad \mathrm{Hess}(J)(x) = 2A^*A \in \mathscr{S}_n^+(\mathbb{R}) \text{ (resp. } \mathscr{H}_n^+(\mathbb{C})).$$

Ainsi, J est une fonction convexe et donc x minimise J si et seulement si x est un point critique de J, c'est-à-dire que x est solution de $Ax = b$ au sens des moindres carrés si et seulement si :

$$\nabla J(x) = 0 \quad \text{i.e.} \quad A^*Ax = A^*b.$$

Étape 2 : le vecteur $A^\dagger b$ est solution de l'équation normale
Si $A = U\Sigma V^*$, alors :

$$A^*AA^\dagger b = V\Sigma^T\Sigma V^*V\Sigma^\dagger U^*b = V\Sigma^T\Sigma\Sigma^\dagger U^*b = V\Sigma^T \begin{pmatrix} I_r & \mathbf{O}_{r,m-r} \\ \mathbf{O}_{m-r,r} & \mathbf{O}_{m-r,m-r} \end{pmatrix} U^*b = V\Sigma^T U^*b = A^*b.$$

Ainsi, $x_b := A^\dagger b$ vérifie l'équation normale, donc c'est une solution de $Ax = b$ au sens des moindres carrés !

Étape 3 : x_b est l'unique solution de norme minimale
Soit $x \in \mathbb{K}^n$ solution de $Ax = b$ au sens des moindres carrés. On a donc :

$$x - x_b \in \ker\left(A^*A\right).$$

Or, $x_b = A^\dagger b \in \ker\left(A^*A\right)^\perp$. En effet, on a $\ker\left(A^*A\right) = \mathrm{Vect}(v_{r+1}, \ldots, v_n)$ en réutilisant les notations de la preuve de la SVD. Ainsi, il suffit de montrer que x_b est orthogonal à v_i pour $i \in [\![r+1, n]\!]$:

$$\forall i \in [\![r+1, n]\!], \quad \langle v_i, x_b\rangle = \left\langle v_i, V\Sigma^\dagger U^*b\right\rangle = \left\langle U\left(\Sigma^\dagger\right)^T V^*v_i, b\right\rangle = \left\langle U\left(\Sigma^\dagger\right)^T e_i, b\right\rangle = 0$$

car la i-ème colonne de $\left(\Sigma^\dagger\right)^T$ est nulle pour $i > r$. Ainsi, on a :

$$\|x\|_2^2 = \|x_b\|_2^2 + \|x - x_b\|_2^2 \geqslant \|x_b\|_2^2$$

avec égalité si et seulement si $x = x_b$, ce qui termine la preuve de ce beau résultat ! $\square$

Remarque 4.3.2. *On a également les résultats suivants :*

1. $AA^\dagger$ *est la matrice de la projection orthogonale sur* $\mathrm{Im}(A)$,
2. $A^\dagger A$ *est la matrice de la projection orthogonale sur* $\mathrm{Im}\left(A^*\right) = \ker(A)^\perp$.

Démonstration. Cela se voit en remarquant que $AA^\dagger$ et $A^\dagger A$ s'écrivent ainsi :

$$AA^\dagger = U \begin{pmatrix} I_r & \mathbf{O}_{r,m-r} \\ \mathbf{O}_{m-r,r} & \mathbf{O}_{m-r,m-r} \end{pmatrix} U^* = \sum_{i=1}^{r} u_i u_i^* \quad \text{et} \quad A^\dagger A = V \begin{pmatrix} I_r & \mathbf{O}_{r,n-r} \\ \mathbf{O}_{n-r,r} & \mathbf{O}_{n-r,n-r} \end{pmatrix} V^* = \sum_{i=1}^{r} v_i v_i^*,$$

où on a noté encore une fois $(u_1, \ldots, u_m)$ les colonnes de U et $(v_1, \ldots, v_n)$ les colonnes de V. Il ne reste donc plus qu'à justifier que :

1. $\mathrm{Im}(A) = \mathrm{Vect}\,(u_1, \ldots, u_r)$, et
2. $\ker(A)^\perp = \mathrm{Vect}\,(v_1, \ldots, v_r)$.

En effet, les vecteurs u_i et v_i formant une famille orthonormale, la matrice dans la base canonique de la projection sur $\mathrm{Vect}(u_1, \ldots, u_r)$ (resp. $\mathrm{Vect}(v_1, \ldots, v_r)$) est bien $\displaystyle\sum_{i=1}^{r} u_i u_i^*$ (resp. $\displaystyle\sum_{i=1}^{r} v_i v_i^*$).

1. Étant donné que $r = \mathrm{rg}(A)$, il suffira de montrer que $\mathrm{Vect}(u_1, \ldots, u_r) \subset \mathrm{Im}(A)$ par dimension. Or, on a :

$$\forall i \in [\![1, r]\!], \quad u_i = \frac{1}{\sigma_i} A v_i$$

   comme on a vu dans la preuve de la SVD ! Ainsi, $u_i \in \mathrm{Im}(A)$ et cela montre le point 1.

2. Au vu du théorème spectral que l'on a appliqué à A^*A dans la preuve de la SVD, on a :

$$\ker(A) = \ker(A^*A) = \mathrm{Vect}(v_{r+1}, \ldots, v_n) = \mathrm{Vect}(v_1, \ldots, v_r)^\perp$$

   étant donné que la base $(v_1, \ldots, v_n)$ de $\mathbb{K}^n$ est orthonormée.

□

Ce point 1. permet de montrer sans passer par l'équation normale que le vecteur $A^\dagger b$ est solution au sens des moindres carrés de $Ax = b$. Mais attention ! Cela ne montre pas qu'il est de norme minimale ! Le travail fait à l'étape 3 reste donc primordial à présenter.

Remarque 4.3.3. *Il est très important de bien savoir justifier que $\ker(A) = \ker(A^*A)$! Rappelons comment cela se montre. On a une inclusion claire : $\ker(A) \subset \ker(A^*A)$. Prenons alors $x \in \ker(A^*A)$ et montrons qu'il appartient à $\ker(A)$:*

$$\|Ax\|^2 = \langle Ax, Ax \rangle = \langle x, \underbrace{A^*Ax}_{=0} \rangle = 0.$$

Ainsi $Ax = 0$, ce qu'on voulait montrer !

Remarque 4.3.4. *Peut-être qu'avec la rédaction du sujet de maths générales 2023, vous pourrez faire rentrer ce développement dans la leçon 158. Je conseille alors de ne faire que la résolution au sens des moindres carrés en justifiant les résultats de la remarque 4.3.2.*